Emily Morgan
UC Davis

Generative and Item-Specific Knowledge in Language Processing

Abstract: The ability to generate novel utterances compositionally using generative knowledge is a hallmark property of human language. At the same time, languages contain non-compositional or idiosyncratic items, such as irregular verbs, idioms, etc. In this talk I ask how and why language achieves a balance between these two systems—generative and item-specific—from both the synchronic and diachronic perspectives.

Specifically, I focus on the case of binomial expressions of the form "X and Y", whose word order preferences (e.g. bread and butter/#butter and bread) are potentially determined by both generative and item-specific knowledge. I show that ordering preferences for these expressions indeed arise in part from violable generative constraints on the phonological, semantic, and lexical properties of the constituent words, but that expressions also have their own idiosyncratic preferences. I argue that both the way these preferences manifest diachronically and the way they are processed synchronically is constrained by the fact that speakers have finite experience with any given expression: in other words, the ability to learn and transmit idiosyncratic preferences for an expression is constrained by how frequently it is used. The finiteness of the input leads to a rational solution in which processing of these expression relies gradiently upon both generative and item-specific knowledge as a function of expression frequency, with lower frequency items primarily recruiting generative knowledge and higher frequency items relying more upon item-specific knowledge. This gradient processing in turn combines with the bottleneck effect of cultural transmission to perpetuate across generations a frequency-dependent balance of compositionality and idiosyncrasy in the language, in which higher frequency expressions are gradiently more idiosyncratic. I provide evidence for this gradient, frequency-dependent trade-off of generativity and item-specificity in both language processing and language structure using behavioral experiments, corpus data, and computational modeling.